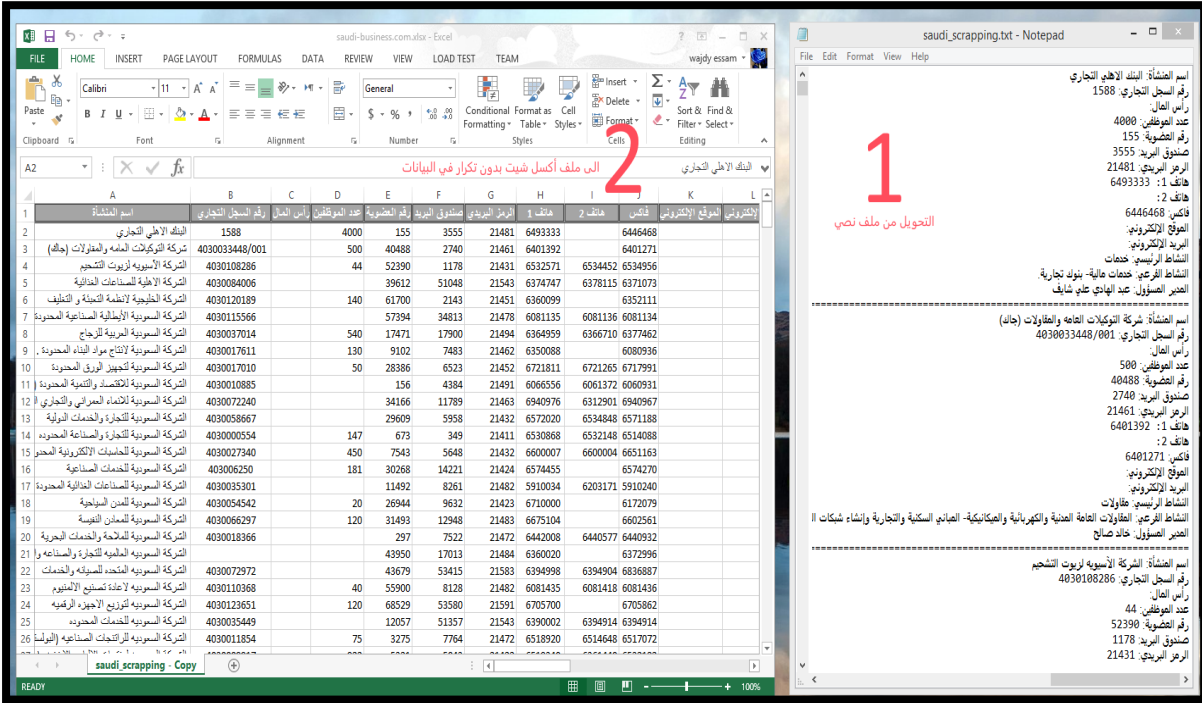


تحويل بيانات من ملف نصي الى اكسل فورمات

بعد تحميل الملف من الرابط <http://bit.ly/1EZrjeV> وجد أن هناك حوالي 500 شركة بمعلومات حولها ولكن طريقة العرض عن طريق الأسطر Row Based والأسهل هو تحويلها للاكسل ك Columns Based حتى يتم البحث والفلتره بها بسهولة.



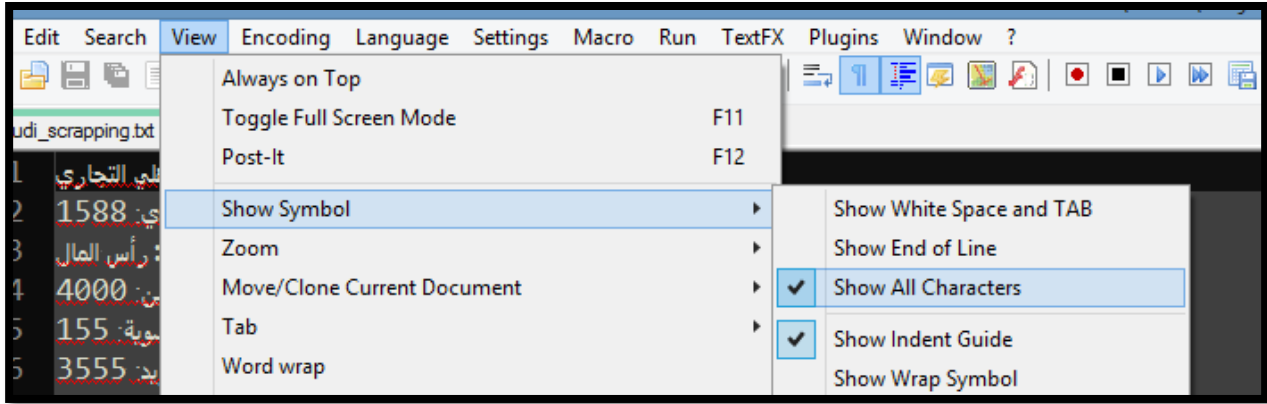
في هذ الملف سوف نوضح الطريقة لتحويل ذلك الملف، ويمكن اتباع هذه الآليه في التحويل لغيرها من الملفات مع الأخذ بالاعتبار ان بعض الخطوات قد تختلف بناء على بنية الملف المراد تحويله.

سوف يتم استخدام برنامج notepad++ في عمليات التحويل هذه، مع العلم بأنه يوجد برامج بديله له سواء في بيئة ويندوز أو لينوكس أو ماك، وبشكل عام اي برنامج محرر نصوص يتوفر فيه خاصية البحث والاستبدال بواسطة Regular أو Extended Mode أو Expression Mode فهو مناسب في الأمر.

نقطة أخرى وهو أن برنامج notepad++ لا يتعامل مع الملفات الضخمة بشكل جيد وقد يهتق Crashing البرنامج اثناء فتح اي ملف كبير أو اثناء القيام بخطوات التحويل، لذلك عند تعاملك مع أي ملف كبير يفضل أن تستخدم البرامج المخصصة للعمل مع الملفات الكبيرة مثل EditPad Lite وهو مجاني في ويندوز.

نبدأ الخطوات على بركة الله:

اولاً: دائماً الأفضل اظهار Symbols الموجودة والمخفية في الملف والتي عادة اغلبها يكون في آخر كل سطر لمعرفة نوعية ال New Lines، فبعد أن تقوم بفتح الملف تقوم عن طريق قائمة View ومن ثم اختيار Show all Characters



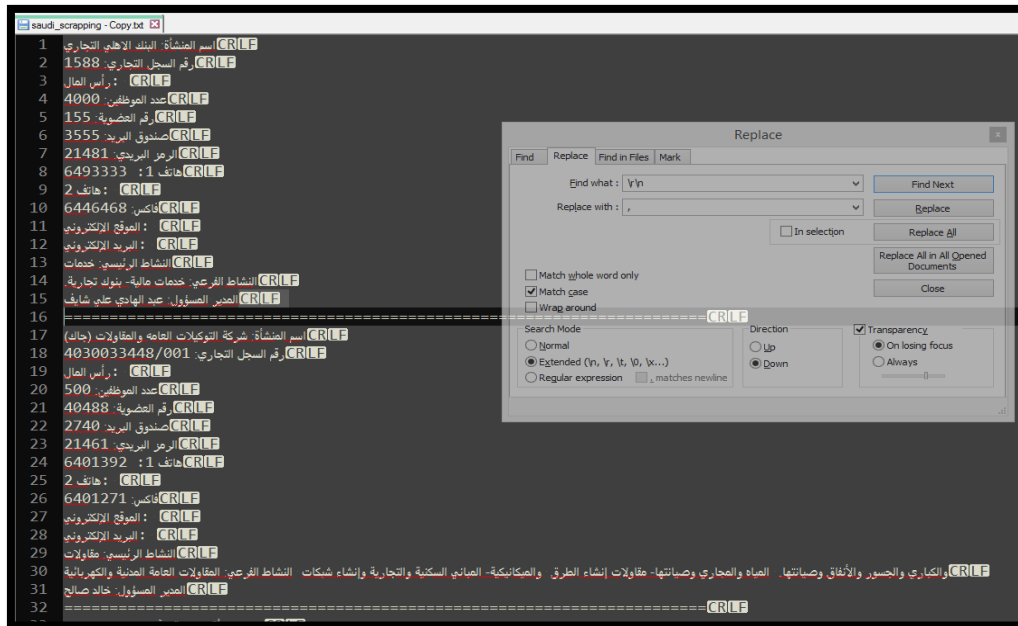
ويمكنك اخفائها عن طريق إزالة الصح على Show all Characters بعدما تعرف نوعية ال New Line، بعد أن قمت باظهارها، سوف تجد الان الرموز الموجودة في آخر كل سطر مثلاً CRLF أو CR أو LF على حسب طريقة ال New Line المستخدمة في الملف.

لماذا نقوم بمشاهدتها؟

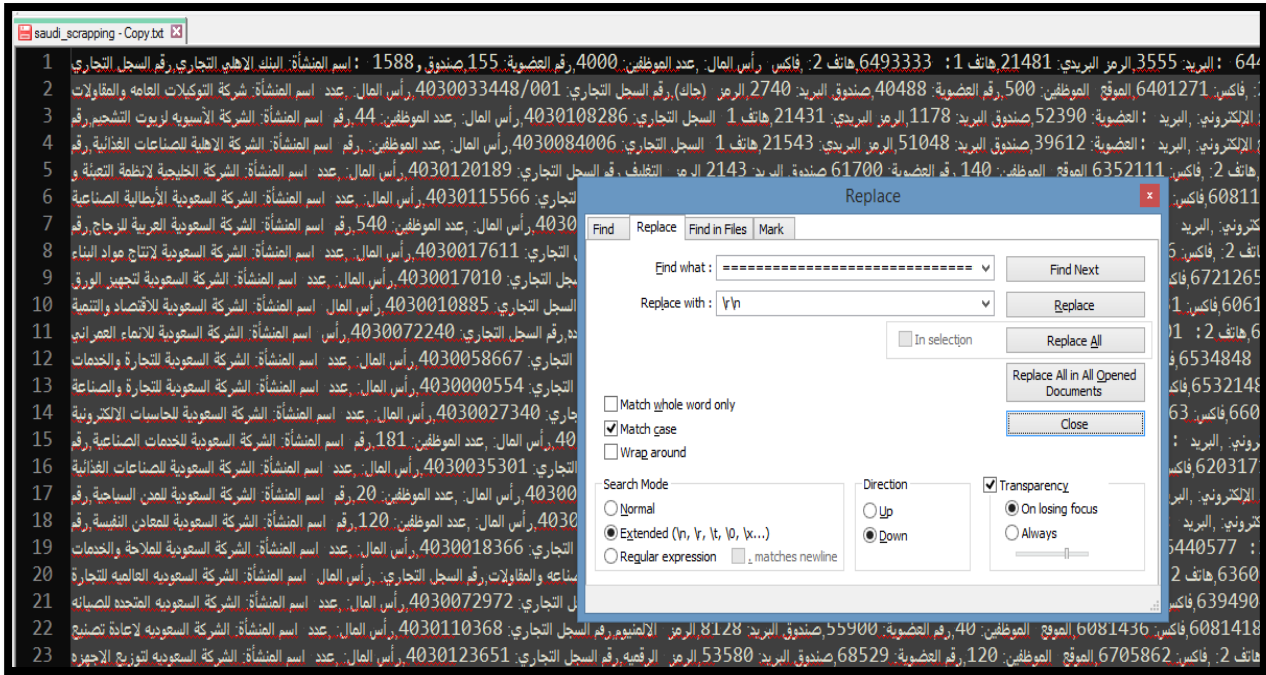
لأنه أحياناً تجد ملفات مستخدم بها أكثر من طريقة New Line في نفس الوقت، وحتى تقوم بعملية التحويل المناسبة عليك أن تعرف نوعية ال New Line المستخدمة، في آخر هذه الملف سوف نضع مثال على ملف آخر به أكثر من New Line مستخدمة بطريقتي ويندوز ولينوكس ونعرف كيف يتم اصلاح ذلك من خلال ال notepad++.

ثانياً: بما أن برنامج الأكسل يستطيع قراءة ملفات CSV واختصاراً Comma Separated Value وهي تعني أن هناك معامل الفاصلة يفصل بين كل معلومة وأخرى، فنحن نحتاج أن نحول الملف لذلك الفورمات.

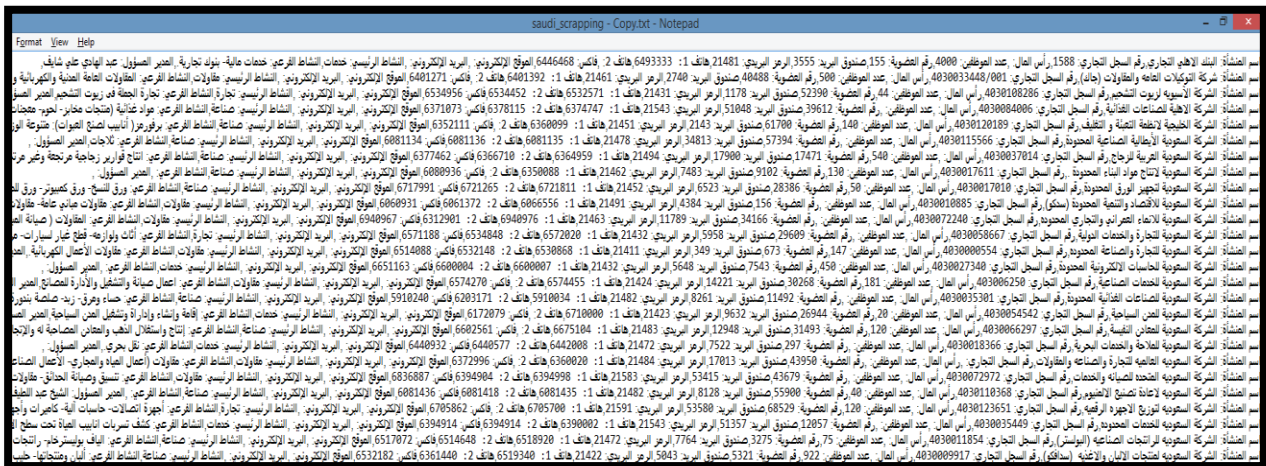
لذلك سوف نقوم بتحويل كل الأسطر New Lines الى فاصلة، وبالتالي جميع البيانات تكون في سطر واحد وذلك عن طريق عمل Find And Replace من خلال CTRL+F والذهاب للReplace ومن ثم اختيار Extended Mode والذي يسمح بالبحث من خلال \n,\r,\t وغيرها، وسوف تكون جملة البحث هي \r\n (أي CRLF) ويتم استبدالها بالفاصلة كما في الصورة التالية، وتقوم بعمل Replace All لتغيير كل الأحرف التي طابقت جملة البحث في هذا الملف.



ثالثاً: بعد عملية التحويل سوف تجد أن جميع البيانات هي في سطر واحد، وسوف نستفيد من علامة ===== حتى نفرق بين كل بيانات شركة على حدة ، لذلك سوف نقوم بعمل فصل بهذه العلامة واستبدالها بسطر جديد من خلال البحث عن ===== وتبديلها ب \r\n وعمل Replace All كما في الصورة التالية بعد تطبيق زر التبديل.

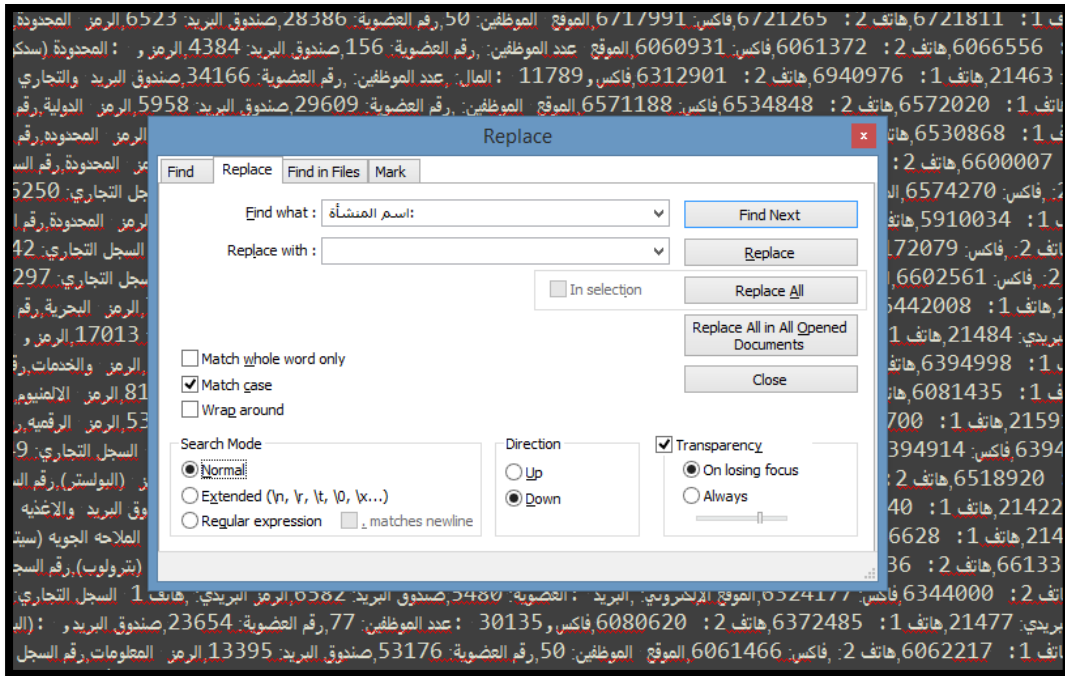


وهذا هو الملف أيضاً تم فتحه عن طريق النوت باد بوضع ال Right To Left (بعد فتحه بالنوت باد نقوم ب CTRL+SHIFT التي على اليمين)، وايضاً تستطيع عمل ال Right To Left في ال Notepad++

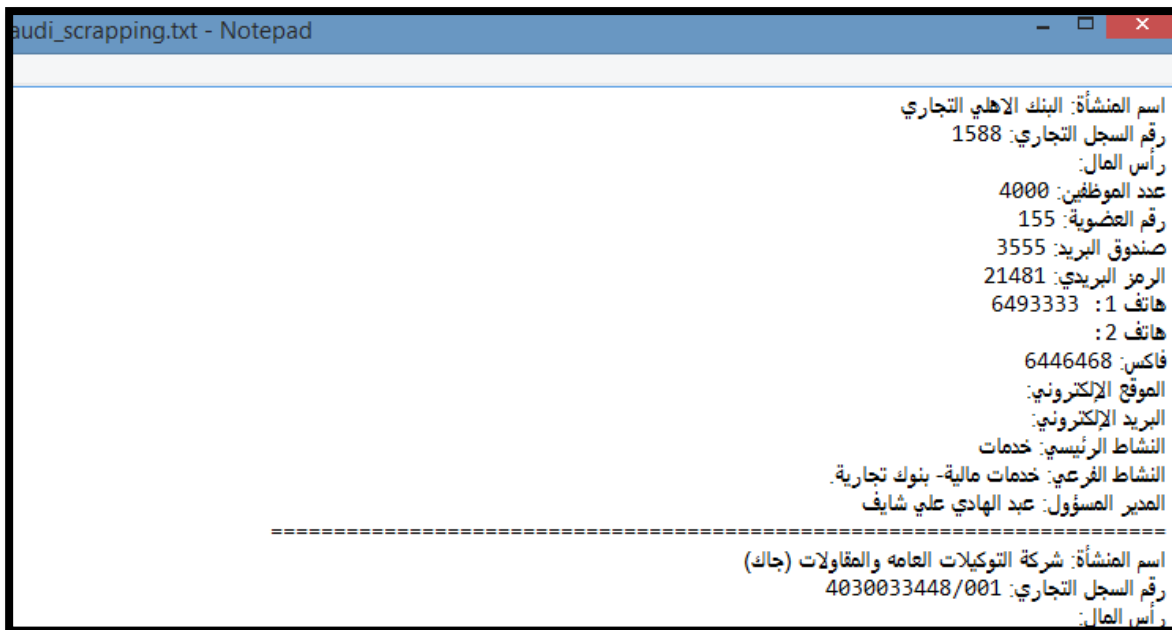


كما هو واضح الملف أصبح مرتب الان وكل شركة في سطر، ونريد تحويله الى صيغة اكسل.

رابعاً: قبل التحويل سوف نحتاج لمسح البيانات المكررة في كل سطر (اسم المنشأة، رقم السجل التجاري والخ) لأنها سوف تكون كإسم للمعمود في الأكسل ولا توجد حاجة لكتابتها في كل سطر، لذلك نقوم بمسح تلك الكلمات من خلال أيضاً البحث عن كل اسم ثابت ومسحه، وطريقة المسح هي عن طريق استبدال النص بنص فارغ كما بالصورة التالية



تم حذف اسم المنشأة: ولاحظ وجود : لأنها داخله في النص الثابت في الملف، وسوف نكرر هذه الخطوة لجميع الأسماء الثابتة في الملف، اسم المنشأة: رقم السجل التجاري: ، الخ والتي تتواجد في بداية أي سطر في الصورة الأولى قبل التحويل:



خامساً: فقط للتأكد من ال Encoding وأنه UTF-8 (والتي تendl على وجود العلامة BOM) فهذه العلامة BOM ضرورية لملفات ال Office حتى تستطيع قرائه ال encoding بشكل صحيح، قم بفتح الملف بال++Notepad والنظر ال Encoding وأنه UTF-8 وغيرها اذا كانت مثلاً UTF-8 Without BOM ال UTF-8.

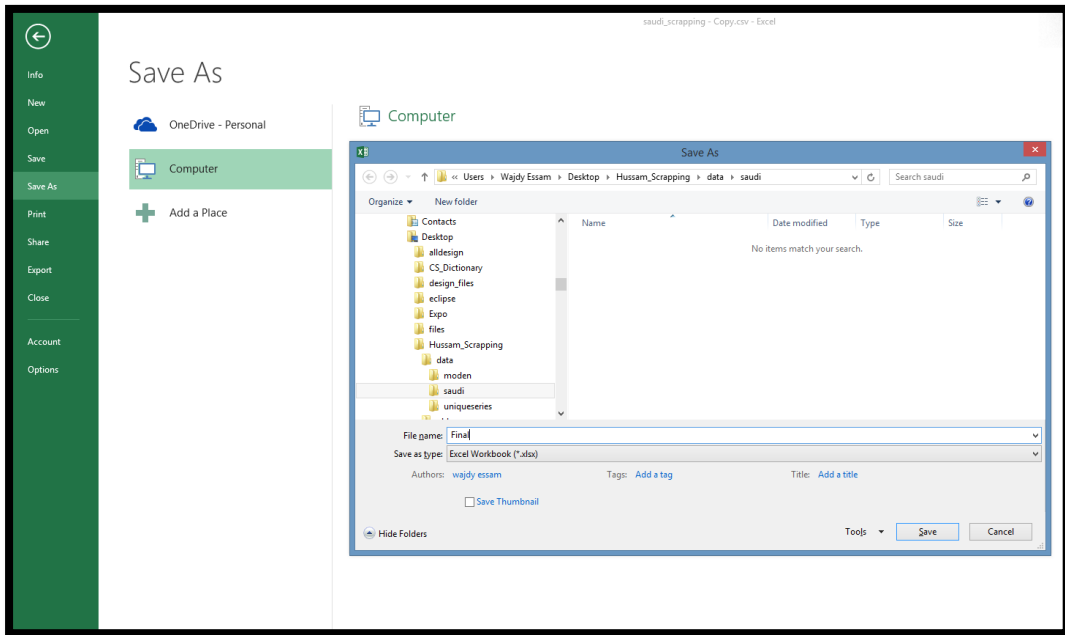
سادساً: نقوم بتحويل امتداد الملف الى CSV عن طريق عمل Rename للملف وضع الامتداد csv، ومن ثم نقوم بفتحه بالاكسل بشكل عادي Double Click لأنه يقبل فتح ملفات CSV

Name	Date modified	Type	Size
saudi_scrapping - Copy.csv	5/14/2015 11:10 AM	Microsoft Excel C...	268 KB

الملف بعد فتحة بالاكسل سوف يكون بهذا الشكل، والأفضل حفظه بصيغة الأكسل حتى يسهل العمل عليه في الأكسل فيما بعد.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
البنك الاهلي	1588		4000	155	3555	21481	6493333		6446468				خدمات مالية- بنوك	عبد الهادي
شركة التوكيز	4030033448/O		500	40488	2740	21461	6401392		6401271				المقاولات العامة المد	خالد صالح
الشركة الاسر	4030108286		44	52390	1178	21431	6532571	6534452	6534956				تجارة الجملة في زيب	طارق عبد
الشركة الاط	4030084006			39612	51048	21543	6374747	6378115	6371073				مواد غذائية (منتجات	صناعة
الشركة الخبز	4030120189		140	61700	2143	21451	6360099		6352111				برفورمز (أنابيب لم	محمود المهدي
الشركة السم	4030115566			57394	34813	21478	6081135	6081136	6081134				صناعة	تلاجات
الشركة السم	4030037014		540	17471	17900	21494	6364959	6366710	6377462				صناعة	د/طارق عت
الشركة السم	4030017611		130	9102	7483	21462	6350088		6080936				صناعة	
الشركة السم	4030017010		50	28386	6523	21452	6721811	6721265	6717991				صناعة	أحمد نبيل الأ
الشركة السم	4030010885			156	4384	21491	6066556	6061372	6060931				صناعة	
الشركة السم	4030072240			34166	11789	21463	6940976	6312901	6940967				صناعة	مقاولات مبانى عامة
الشركة السم	4030058667			29609	5958	21432	6572020	6534848	6571188				صناعة	مقاولات (صيانة ا
الشركة السم	4030000554		147	673	349	21411	6530868	6532148	6514088				صناعة	أثاث ولوازمه- قطع
الشركة السم	4030027340		450	7543	5648	21432	6600007	6600004	6651163				صناعة	مقاولات الأعمال الك
الشركة السم	403006250		181	30268	14221	21424	6574455		6574270				صناعة	خدمات
الشركة السم	4030035301			11492	8261	21482	5910034	6203171	5910240				صناعة	م.صالح احمد
الشركة السم	4030054542		20	26944	9632	21423	6710000		6172079				صناعة	اصعمال صيانة والتش
الشركة السم	4030066297		120	31493	12948	21483	6675104		6602561				صناعة	حساء ومرق- زيد- م
الشركة السم	4030018366			297	7522	21472	6442008	6440577	6440932				صناعة	إقامة وإنشاء وإدارة
الشركة السم	4030072972		43950	17013	21484	6360020		6372996	6372996				صناعة	م / محمد هاد
الشركة السم	4030072972		43679	53415	21583	6394998	6394904	6836887	6836887				صناعة	مقاولات (أعمال الم
الشركة السم	4030110368		40	55900	8128	21482	6081435	6081418	6081436				صناعة	مقاولات
الشركة السم	4030123651		120	68529	53580	21591	6705700		6705862				صناعة	تنسيق وصيانة الحد
الشركة السم	4030035449			12057	51357	21543	6390002	6394914	6394914				صناعة	الشيخ عبد الأ
الشركة السم	4030011854		75	3275	7764	21472	6518920	6514648	6517072				صناعة	عبد الرحمن

سابعاً: حفظ الملف بصيغة أكسل xlxs من خلال File->Save As حتى يتحول الملف لصيغة الأكسل



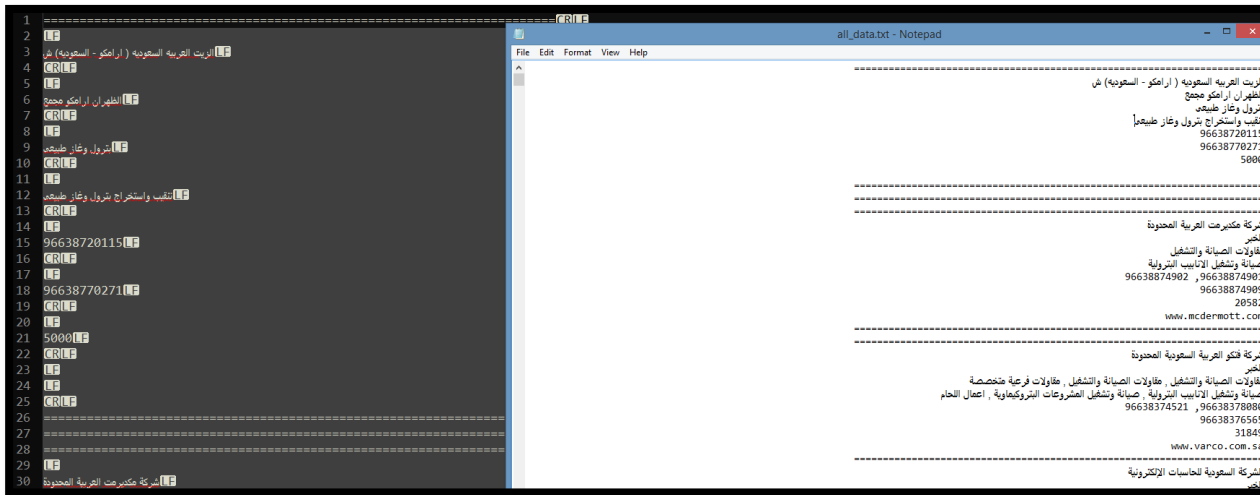
يمكنك اخيراً إضافة السطر الأول وإضافة أسماء الأعمدة وتعديل لونها قليلاً بما تشاء

السطر رقم	الاسم الرئيسي	العنوان الإلكتروني	الموقع الإلكتروني	الفاكس	هاتف 2	هاتف 1	الرمز البريدي	صندوق البريد	رقم العضوية	عدد الموظفين	رقم السجل التجاري	اسم المنشأة
1	خدمات			6446468	6493333	21481	3555	155	4000	1588		البنك الأهلي التجاري
2	مقاولات			6401271	6401392	21461	2740	40488	500	4030033448/001		شركة التوكيلات المعمه والمقاولات (مكاف)
3	تجارة			6534956	6532571	21431	1178	52390	44	4030108286		الشركة الايطيه لزيتون التجميم
4	صناعة			6371073	6374747	21543	51048	39612		4030084006		الشركة الايطيه للمساعدات الغذائية
5	صناعة			6352111	6360099	21451	2143	61700	140	4030120189		الشركة الخديجية لانظمة التناطة و التنظيف
6	صناعة			6081134	6081135	21478	34813	57394		4030115566		الشركة السعودية الايطية للصناعة المحدودة
7	صناعة			6377462	6364959	21494	17900	17471	540	4030037014		الشركة السعودية العربية للتراجيح
8	صناعة			6080936	6350088	21462	7483	9102	130	4030017611		الشركة السعودية لاناج مواد البناء المحدودة .
9	صناعة			6717991	6721811	21452	6523	28386	50	4030017010		الشركة السعودية لتمويل القروض المحدودة
10	مقاولات			6060931	6061372	21491	4384	156		4030010885		الشركة السعودية للاقتصاد والتنمية المحدودة (سكرا)
11	مقاولات			6940967	6940976	21463	11789	34166		4030072340		الشركة السعودية لخدمات العمالي والتجاري المحدودة
12	تجارة			6571188	6534848	21432	5958	29609		4030058667		الشركة السعودية للتجارة والخدمات الدولية
13	مقاولات			6514088	6532148	21411	349	673	147	4030005554		الشركة السعودية للتجارة والصناعة المحدودة
14	خدمات			6651163	6600004	21432	5648	7543	450	4030027340		الشركة السعودية لخدمات الالكترونية المحدودة
15	مقاولات			6574270	6574455	21424	14221	30268	181	4030062590		الشركة السعودية للخدمات الصناعية
16	صناعة			5912029	5910034	21482	8261	11492		4030035301		الشركة السعودية للمساعدات الغذائية المحدودة
17	خدمات			6602561	6710000	21423	9632	26944	20	4030054542		الشركة السعودية للنسج النسيجية
18	صناعة			6440932	6442008	21472	7522	297	120	4030066297		الشركة السعودية للخدمات النسيجية
19	خدمات			6372996	6360020	21484	17013	43950		4030018366		الشركة السعودية للخدمات البحرية
20	مقاولات			6838887	6394904	21583	53415	43679		4030072972		الشركة السعودية للمعدات الثقيلة والصناعة والمقاولات
21	صناعة			6081436	6081435	21482	8128	55900	40	4030110368		الشركة السعودية لاجهزة تصنيع الانابيب
22	تجارة			6705862	6705700	21591	53580	68529	120	4030123651		الشركة السعودية لاجهزة تصنيع الانابيب

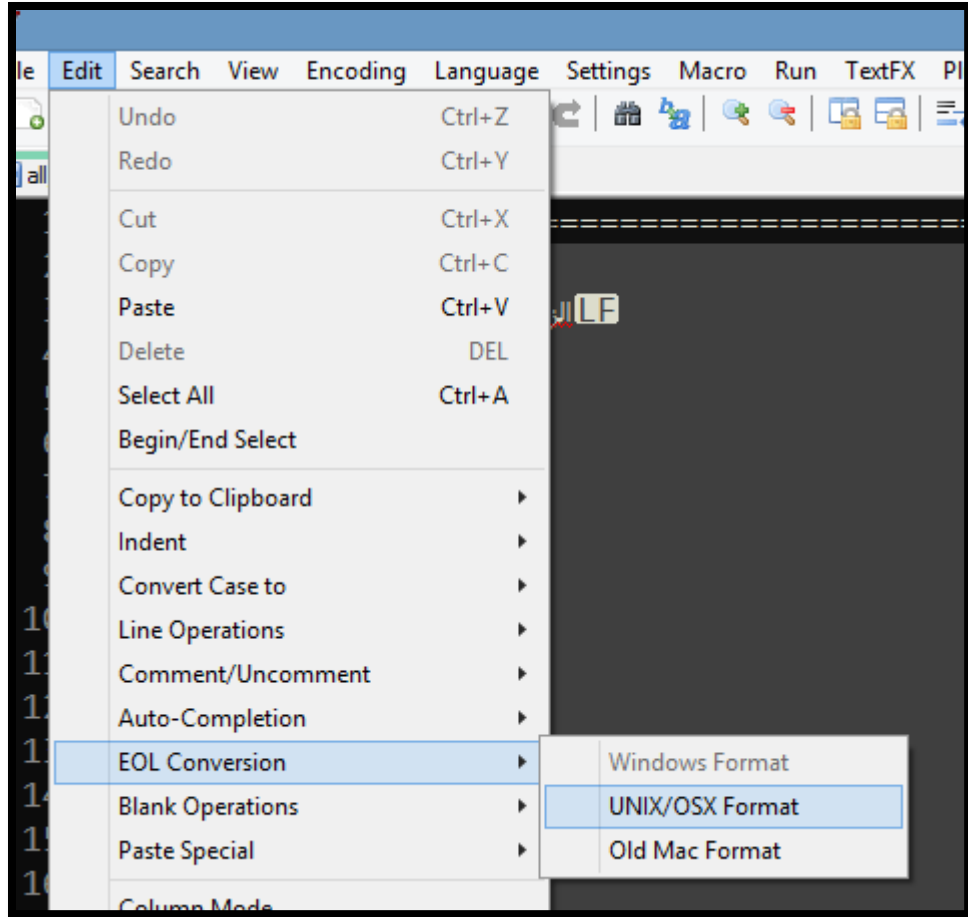
وهكذا تكون قد تمت عملية تحويل الملف بنجاح! وتستطيع بعدها بسهولة تطبيق اي Filter من خلال الأكسل حتى تقلتر الأسطر وتظهر من تريد بالضغظ على الهيدر في العمود الذي تريده واختر Filter By Value وستجد القائمة المنسدلة وحينها اختر القيمة التي تريد وستجد أن الأسطر قم تمت فلترتها بناء على ما اخترته.

لماذا تحتاج لأن ترى كل الأحرف احياناً في الملف؟

هذا مثال لأحد الملفات فيه ايضاً مجموعه من البيانات وكان المطلوب هو تحويلها الى Excel، وعندما تفتح الملف بالنوت الباد العادي سوف تجد الصورة على اليمين ، ولكن عندما تفتحه بالنوت باد بلس بلس سوف تجد على اليسار ، ولاحظ الفرق بأن LF لا يعرضه Notepad كمسافة بعكس ال Notepad++ والذي يعرض ال LF بأنها مسافة (طبعاً مع تفعيل خاصية Show all Characters) التي سبق توضيحها



قبل أن نتعامل مع هذا الملف بالطريقة السابقة اعلاه، يجب أن يتم اعتماد New Lines واحدة في الملف وجعل الملف كله اما LF أو CRLF وال أفضل في ويندوز، والأمر يتم كالتالي عن طريق تحويلها جميعها الى Unix Format ومن ثم ارجاعها الى Windows Format

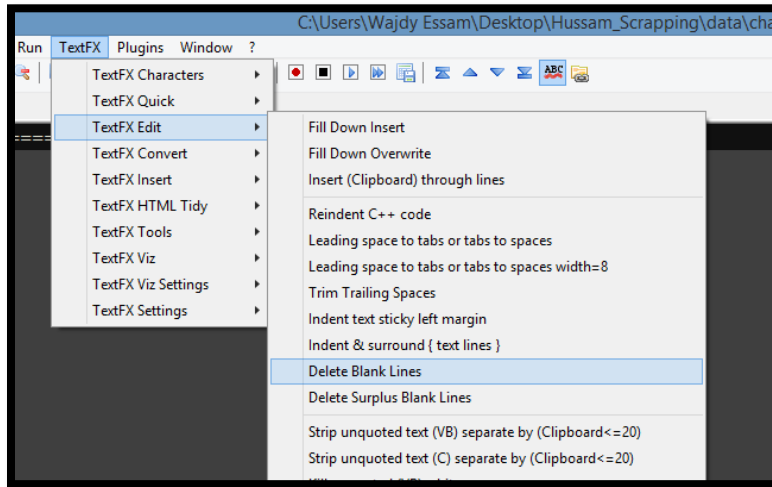


بعد ذلك سوف تتحول كلها الى LF قم بإعادة الخطوات مجدداً واختر Windows Format وهكذا جميعها سوف تتحول الى CRLF



الخطوة التالية هي حذف كل الاسطر الفارغه ويمكن استخدام خاصية البحث والتبديل ، ولكن كطريقة اخرى فأن ال-Notepad++ فيها مجموعه من الاضافات الرائعه التي تقوم بالعديد من المهام الاضافية لمحرر النصوص، وأحد هذه الاضافات هي TextFX حيث فيها

العديد من المهام التي تفيد اثناء العمل على الملفات النصية، وأحدها هو أمر يقوم بحذف كل الأسطر الفارغه، فقط قم بتظليل كامل الملف CTRL+A واختر الخيار بعد تحميل الاضافة بالطبع



الان تحول الملف ورجع للشكل الصحيح

```
all_data - Copy.txt
1 =====CRLF
2 الزيت العربي السعودي ( ارامكو - السعودية ) شCRLF
3 الظهران ارامكو مجمعCRLF
4 بترول وغاز طبيعيCRLF
5 تنقيب واستخراج بترول وغاز طبيعيCRLF
6 96638720115CRLF
7 96638770271CRLF
8 5000CRLF
9 =====CRLF
0 =====CRLF
1 =====CRLF
2 شركة مكعب مت العربية المحدودةCRLF
```

أحياناً، قد يكون في الملف فيه فاصلة , بين البيانات ، فقبل أن تحول الملف الى CSV عليك بحذف هذه الفاصلة (والا ستخرب عليك البيانات بعد تحويلها حيث ستفصل البيانات الأصلية عن بعضها) فيجب ان تقوم بعمل بحث عن , واستبدالها بأي شيء آخر وليكن مثلاً علامة الشرطة - ، وهكذا تكون قد عملت Cleanup للبيانات وجهازتها حتى تتحول الى CSV بعد ذلك تستطيع القيام بنفس الطريقة الموضحة في هذا الملف في تحويل البيانات ، وذلك عن طريق استبدال ال CRLF بالفاصلة وبعدها تبديل ال == بالCRLF ومن ثم حفظ الملف CSV وفتحه بالأكسل وحفظه بصيغة الأكسل.

وتشكراً لكم

اي سؤال يمكنكم طرحه في صفحة انفورماتيك، أو عبر بريدي الموضحة أدناه

<https://www.facebook.com/informatics.arab>

wajdyessam@hotmail.com