Untangler®
How it works

Untangler® is focussed on one thing - automatically recognising and transforming inconsistent tabular data (spreadsheets, lists, schedules of assets, etc.) from multiple sources to your required standard.

It uses cognitive automation, emulating human behaviour, to take decisions on its own. Unlike other systems, there is no need to open and prepare the inbound data files.

Untangler® does 3 things:

1. Finds relevant data
2. Recognises and validates data
3. Transforms data for output

Making Untangler® work for you
We work with you to define the nature of the data needed. We help set up a custom output template to map the untangled data to your required output format.

A library of data dictionaries is used as the single source of knowledge to identify data from the input. Each data dictionary contains definitions that include aliases, type information and constraints. Your data dictionaries are based on your specific business requirements.

A test suite of sample files is used to tune the performance to an ever-increasing level of accuracy.

Untangler® carries out a range of automated data recognition and transformation strategies on each input data file.

It starts by running block processing in order to identify each relevant block of data and its relationship to others. Redundant rows are marked for ignoring and hidden columns can be configured for exclusion.

Column headings are then decomposed into various elements (potential data dictionary entries, punctuation, entities, temporal qualifiers, etc.). A confidence score is calculated for each and those that are high enough become candidates for further processing.

Having determined a set of possible column candidates, they are checked by type handlers. These validate (and optionally correct) data cells for specific types and their constraints. A type is an attribute in a data dictionary and includes number, string, money, enumeration (a complete list), date, etc.

The combination of the confidence scores from column decomposition and column data validation are used to select the best input column for the desired output template. The output file can be generated in a number of formats, including MS Excel and CSV. An API is available if you'd prefer to integrate the service into your own systems.

Untangl

hello@untangl.co.uk
0207 043 2550
untangl.co.uk/untangler

Untangler® example

Untangler® detects *blocks* of data and seeks to establish which is the most likely match, using pattern recognition based on on the output specification.

Company Name	ABC Partners Ltd
Company Postcode	LS13 4RT
Company address	The Business Centre, Holden Road, Leeds
Company phone number	0113 334 55667

Surname	Forename	Gender	Date of Birth	Occupation	Basic Annual Salary	Workplace Postcode
Smith	Kate	F	01/01/1972	MD	£150,000	TW11 7RT
Jones	Sophie	F	1.1.80	Sales	£50,000	TW117RT
Hughes	Jane	female	02/03/1987	Manager	25000.00	Teddington TW11 7RT
				total	£225,000	

Once the block has been detected, Untangler® works on data recognition and validation. The primary aim is to determine columns by matching headings and column contents against the output requirement. Column headings are cross checked against data dictionaries, e.g. "Surname" to last_name. It also validates the column data. In this example, dates of birth are checked against a specified date range and 2 digit years converted to YYYY.

2	1	4	3	x	6	5

Surname	Forename	Gender	Date of Birth	Occupation	Basic Annual Salary	Workplace Postcode
Smith	Kate	F	01/01/1972	MD	£150,000	TW11 7RT
Jones	Sophie	F	1.1.80	Sales	£50,000	TW117RT
Hughes	Jane	female	02/03/1987	Manager	25000.00	Teddington TW11 7RT

Aaa	Aaa	M / F	DD/MM/YYYY	1,000.00	AANN NAA

The final task is to transform the data to the output specification. In this example, the following changes have been made: columns renamed and reordered, fields formatted to required standard (gender, postcodes, salaries, dates of birth). Custom functions can also be applied - in this example, incremental ID row numbers are appended.

ID	First Name	Last Name	Date of Birth	Gender	Postcode	Salary
1	Kate	Smith	01/01/1972	F	TW11 7RT	150,000.00
2	Sophie	Jones	01/01/1980	F	TW11 7RT	50,000.00
3	Jane	Hughes	02/03/1987	F	TW11 7RT	25,000.00

Untangl

hello@untangl.co.uk

0207 043 2550

untangl.co.uk/untangler