Source text (we are the translator/interpreter)

P. R. O. Bally (1959) "Monadenium arborescens". Candollea 17:25-26.

Coming from Tanzania, this is a <u>robust growing species</u> and is a <u>semi woody succulent</u>, forming a <u>lightly branched</u> shrub/tree up to 4.25 metres high. The <u>stems can grow</u> to 10 cm. <u>thick</u>, are <u>five angled</u> and may be slightly <u>spirally twisted</u>.

They are erect and may be solitary or in twos. If branched, the branches are quite slender, grow erect, and are some 30-60 cm. apart. They are smooth, and covered in a green bloom. Leaf scars, which are 10 mm. in diameter, are borne 4-7 cm. apart and below each leaf scar is a small tubercle which on older plants has a small reddish/brown spine, but a more robust one up to 2 cm. long on is produced on younger plants. The leaves are crowded terminally around the ends of the stems and are produced from the angles of the stems. They are obovate, pointed and heart shaped, 7-19 cm long and 5.6-11 cm. wide. Flowering takes place from an eye situated directly above the leaf scar and several cymes may be produced near the apex of the branches with peduncles 6-7 cm. long and 5-6 mm. thick. The colour of the inflorescence is red. This species is not in general cultivation due to its rapid growth and size.

The process for manual corpus construction

- We want to build a bilingual specialised comparable corpus for the translation task (English → Italian)
- Two stages:
 - a) Source language corpus component (English)
 - b) Target language corpus component (Italian)

Searching for similar SL (English) texts for the corpus

- We look for:
 - web pages in English, as similar to our ST as possible
 - e.g. searching for 'monadenium' on google.co.uk
- We find, e.g.:

en.wikipedia.org/wiki/Monadenium arborescens

www.sdcss.com/monadenium.html

davesgarden.com/guides/pf/go/65135/

www.gardening.eu/plants/Succulent-Plants/Monadenium-guentheri/3708/

• You can add to the search string: monadenium filetype:pdf

In general pages in pdf format tend to be more informative and authoritative





Uninformative, different genre

SDCSS Homepage

k to Articles

The Genus Monadenium

by Tom De Merritt (Espinas Y Flores Oct. 2000)

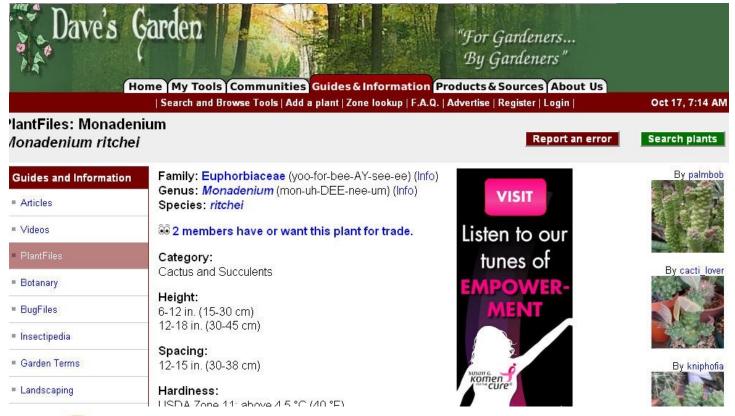
Monadenium belongs in the Euphorbiaceae family, which contains about 320 genera and over 8000 species of geographically widespread plants, which range from annual herbs to large trees. This genus Monadenium has 50 known species, 46 of which are confined to tropical East Africa. Its eastern distribution line is the Indian Ocean. This genus does not exist in Madagascar. Throughout its area of distribution endemic species, varieties and forms of monadenium are numerous, with their habitat often being restricted to a singular hilltop. This high rate of endemism suggests a recent origin and/or a strict response to its environment, which sets limits on the expansion of distribution.

All Euphorbiaceae have a milky sap that may be harmful to the touch. The inflorescence has a complicated structure, based on the cyathium, which consists of an involucre containing one reduced female flower and several male flowers. There are protective bracts and nectarines. The fruit is a capsule, which explodes on reaching maturity.

The word "Monadenium" is derived from Greek, meaning "Single Gland". F. Pax originated the name in 1895 based on the initial species of the genus to be described, Monadenium coccineum. This distinguishing characteristic for which the name was applied emanates from the flower structure. Flowers consist of a cup-like bracteate head, truncate at the top, open at one side or below the middle with a "single gland" of varying

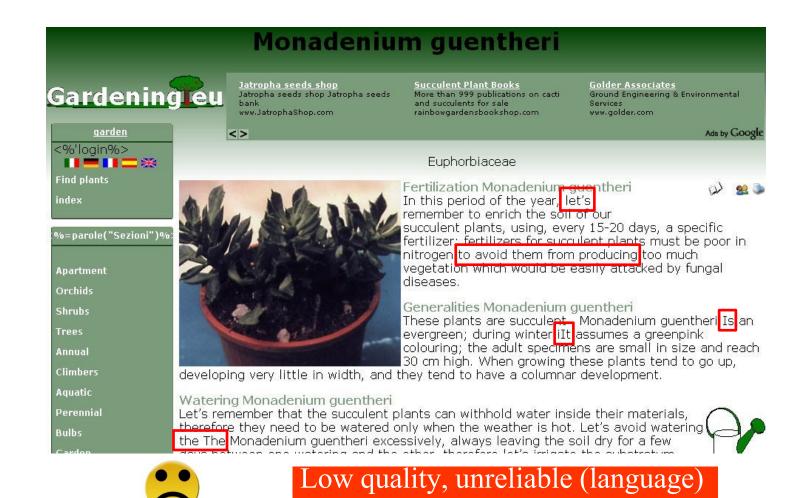


Very informative, authoritative (source: San Diego Cactus Society), similar genre (journal





Uninformative, little connected text, different function (promotional) and genre



Searching for TL texts

- We look for "monadenium" in Italian (reliable) webpages, e.g.:
 - http://www.giardinaggio.it/grasse/singolegrasse/Mo nadenium/Monadenium.asp

Practical considerations: file types

- Corpus files must be downloaded/saved in this format:
 - Simple/pure text (.txt)
 - save as "text only"
- Common formats of online texts
 - HTML
 - File → save as → xxx.txt
 - (just modify the file extension)
 - Microsoft Word
 - Save as → xxx.txt
 - File type → plain text ".txt" → ok (ignore any error message)
 - pdf
 - image/"dead pdf" (not good) vs. searchable pdf (OK)
 - edit → select all → copy → paste in a new text file → save

these must be converted into (saved as) .txt format